Virtual reality measures in neuropsychological assessment: a meta-analytic review

Neguț, A., Matu, S. A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist*, *30*(2), 165-184. doi: 10.1080/13854046.2016.1144793

Virtual reality measures in neuropsychological assessment: a meta-analytic review

Alexandra Neguț, Silviu-Andrei Matu, Florin Alin Sava, and Daniel David

Alexandra Neguț, Evidence-based Assessment and Psychological Interventions Doctoral School; The International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeș-Bolyai University, No. 37, Republicii Street 400015, Cluj-Napoca, Cluj, Romania. Telephone number: 0264434141, e-mail address: alexandra_negut@yahoo.com

Silviu-Andrei Matu, Department of Clinical Psychology and Psychotherapy, Babeș-Bolyai University, No. 37 Republicii Street 400015, Cluj-Napoca, Cluj, Romania. Telephone number: 0264434141, e-mail address: silviu.matu@ubbcluj.ro

Florin Alin Sava, Department of Psychology, West University of Timișoara, No. 4, Vasile Pârvan Bulevard 300223, Timișoara, Timiș, Romania. Telephone number: 0256592252, e-mail address: afsava@gmail.com

Daniel David, Department of Clinical Psychology and Psychotherapy , Babeș-Bolyai University, No. 37 Republicii Street 400015, Cluj-Napoca, Cluj, Romania; Icahn School of Medicine at Mount Sinai, New York. Telephone number: 0264434141, e-mail address: danieldavid@psychology.ro

Correspondence concerning this article should be addressed to: Evidence-based Assessment and Psychological Interventions Doctoral School; The International Institute for the Advanced Studies of Psychotherapy and Applied Mental Health, Babeș-Bolyai University, No. 37, Republicii Street 400015, Cluj-Napoca, Cluj, Romania. Telephone number: 0264434141, e-mail address: alexandra_negut@yahoo.com

Virtual reality measures in neuropsychological assessment: a meta-analytic review

Acknowledgement

“This work was possible due to the financial support of the Sectorial Operational Program for Human Resources Development 2007-2013, co-financed by the European Social Fund, under the project number POSDRU/159/1.5/S/132400 with the title Young successful researchers – professional development in an international and interdisciplinary environment.”

Author Disclosure Statement

No competing financial interests exist.

Virtual reality measures in neuropsychological assessment: a meta-analytic review

Abstract

Objective: Virtual reality-based assessment is a new paradigm for neuropsychological evaluation, that might provide an ecological assessment, compared to paper-and-pencil or computerized neuropsychological assessment. Previous research has focused on the use of virtual reality in neuropsychological assessment, but no meta-analysis focused on the sensitivity of virtual reality-based measures of cognitive processes in measuring cognitive processes in various populations.

Method: We found eighteen studies that compared the cognitive performance between clinical and healthy controls on virtual reality measures.

Results: Based on a random effects model, the results indicated a large effect size in favor of healthy controls (g = .95). For executive functions, memory and visuospatial analysis, subgroup analysis revealed moderate to large effect sizes, with superior performance in the case of healthy controls. Participants' mean age, type of clinical condition, type of exploration within virtual reality environments, and the presence of distractors were significant moderators.

Conclusions: Our findings support the sensitivity of virtual reality-based measures in detecting cognitive impairment. They highlight the possibility of using virtual reality measures for neuropsychological assessment in research applications, as well as in clinical practice.

Keywords: neuropsychological assessment, virtual reality, ecological validity

Virtual reality measures in neuropsychological assessment: a meta-analytic review

Virtual reality systems are a form of interactive and advanced computer technology that generates a 3D environment. They use a human-computer interfaces in a variety of technological tools, such as head-mounted displays (HMDs) for the visual input, trackers and headphones for the acoustic input, video capture systems, data gloves or joysticks in order to enhance the means of interaction (Gamberini, 2000; Ku et al., 2003; Schultheis, Himelstein, & Rizzo, 2002). These generate a computerized representation of the real world, where the person is immersed, and allow the person to navigate or to interact with the virtual world, to see it from different angles and to manipulate it, helping the subject to develop a sense of presence in the virtual world (Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001; Lalonde, Henry, Drouin-Germain, Nolin, & Beauchamp, 2013; Rheingold, 1991).

Virtual reality scenarios are promising tools for neuropsychological assessment (Henry, Joyal, & Nolin, 2012; Parsons, Courtney, & Dawson, 2013; Pugnetti et al., 1998a; Rizzo, Schultheis, Kerns, & Mateer, 2004) and for the rehabilitation of cognitive processes (Chan, Ngai, Leung, & Wong, 2010; Foreman & Stirk, 2005; Rose, Brooks, & Rizzo, 2005). Virtual reality environments have also been tested in the treatment of some psychiatric conditions, such as anxiety disorders (Diaz-Orueta et al., 2012; Opriș et al., 2012; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008).

Virtual reality-based neuropsychological assessment

Neuropsychological assessment is an applied science that focuses on the evaluation of specific activities in the central nervous system (CNS) that are associated with observable behaviors (Lezak, 1995). Neuropsychological evaluation is performed with different types of standardized measurement instruments, that have proved reliability and validity (Morganti, 2004; Schultheis et al., 2002). Paper-and-pencil tests and computerized tests are widely used in neuropsychological assessment and consist of a set of predefined stimuli delivered in a controlled environment via paper-and-pencil or computer systems. They have been found to have a moderate level of ecological validity in predicting real world performance or impairment (Alvarez & Emory, 2006; Chaytor & Schmitter-Edgecombe, 2003; Elkind et al., 2001; Schultheis et al., 2002). Ecological validity refers to the degree to which test results relate to real-life performance (Chaytor & Schmitter-Edgecombe, 2003; Wasserman & Bracken, 2003). Therefore, there is the need to develop neuropsychological tests that evaluate the subject in situations as close as possible to real life and not in a laboratory environment (Chaytor & Schmitter-Edgecombe, 2003). Subsequently, in an attempt to increase the ecological level, new instruments that embed everyday cognitive tasks have been developed. Such tasks describe daily activities like remembering the names of different faces presented in photographs or of the location of various objects, planning a route, solving a practical task, looking at a map and searching for symbols, listening to winning lottery numbers on an audio tape or purchasing specific items. For instance, the Behavioral Assessment of the Dysexecutive Syndrome (Wilson et al., 1986) or Multiple Errands Test (Shallice, & Burgess, 1991) are used for assessing executive functions, the Test of Everyday Attention (Robertson et al., 1996) for attention, while the Rivermead Behavioral Memory Test as a measure of memory (Wilson et al., 1985).

A different approach, with a potentially increased level of ecological validity, is that of using virtual reality to assess cognitive process. Virtual reality-based neuropsychological assessment recreates a real environment in which participants have to solve specific cognitive tasks and their performance is measured within the virtual environment. Based on the conceptual delimitation of ecological validity, that emphasizes the need of similarity between test demands and real life demands, we consider that virtual reality might have an increased level of ecological validity compared to paper-and-pencil or computerized measures. By the use of computers and virtual reality devices, the person has a 3D 360° ―first person‖ view of the scenario, where he can navigate and explore freely. In the virtual scenario, the participant solves cognitive tasks associated with specific cognitive functions. The virtual environments resemble real environments and replicate the challenges found in day to day situations, while maintaining standardized protocols. Various virtual environments have been developed such as: a virtual classroom (Díaz-Orueta et al., 2014; Iriarte et al., 2012; Rizzo et al., 2000; 2006), a virtual seminar room or office (Mania & Chalmers, 2001; Matheis et al., 2007), and a virtual mall (Rand et al., 2007). Virtual reality instruments are currently used for the assessment of executive functions, attention, and impulsivity, cognitive and motor inhibition (Adams et al., 2009; Bioulac et al., 2012; Díaz-Orueta et al., 2014; Elkind et al., 2001; Iriarte et al., 2012; Ku et al., 2003; Parsons et al., 2007), memory and learning (Banville et al., 2010; Matheis et al., 2007; Pugnetti et al., 1998), and visuospatial neglect (Broeren, Samuelsson, Stibrant- Sunnerhagen, Blomstrand, & Rydmark, 2007). Based on these unique features, virtual reality-based assessment tools might have an increased potential to predict everyday functioning (Pugnetti et al., 1999; Pugnetti, Mendozzi, Barbieri, & Motta, 1998b; Rose et al., 2005; Schultheis et al., 2002). Results from these studies point out the potential diagnostic utility of virtual reality tests in neuropsychological assessment, because they could discriminate between healthy and clinical populations.

However, if we take into account the magnitude of the difference between the performance by clinical populations (e.g. patients with ADHD, brain injury, schizophrenia) compared to healthy controls there is a high variability in results (Boiulac et al., 2012; Banville et al., 2010; Ku et al., 2004; Parsons et al., 2007). A similar pattern emerges if we consider the type of cognitive process assessed with the virtual reality-based assessment tool. For instance, for executive functions in case of ADHD children the magnitude of the difference varies from small to large (Bioulac et al., 2012; Parsons et al., 2007). Similarly, in subjects with brain injury the magnitude of the difference for memory assessment is heterogeneous with medium to large effect sizes (Banville et al., 2010; Matheis et al., 2007). Therefore, it would be appropriate to explore in a meta-analysis the variability of these results focusing on the type of cognitive process assessed or clinical population.

Overview of the current study

Although there is evidence in favor of the use of virtual reality measures in neuropsychological assessment (Elkind, 1998; Myers & Bierig, 2000; Riva, 1998; Rizzo et al., 1999), no systematic review has been conducted on this topic. Systematic reviews and meta-analyses have several advantages over narrative ones. First of all, they neutralize selective bias of studies by identifying and synthesizing data from all available studies on a specific topic of research. Meta-analyses also provide an estimate of effect magnitude by combining quantitative data from selected studies. The effect sizes obtained from meta-analyses help to develop a single conclusion with a greater statistical power compared to single studies (Borenstein, Hedges, Higgins, & Rothstein, 2009).

The current meta-analysis aimed: 1) to investigate the sensitivity[1] of virtual reality-based measures of cognitive processes between clinical and healthy populations; 2) to investigate potential moderators of the results.

---

[1] When addressing the diagnostic validity of a neuropsychological test one important aspect is the classification accuracy of the measure (Lezak, 1995). Classification accuracy refers to the correct percentage of correctly identified cases as belonging to either clinical group or healthy control group (Lezak, 1995). This classification accuracy is expressed via indexes like sensitivity/specificity. One approach to establish the sensitivity of a neuropsychological test is by

Potential theoretical moderator variables

Demographic variables

The participants' mean age and the percentage of male participants were considered as potential moderators. Children and young adults may be more familiarized and attracted to computers and technology compared to adults and this can influence the strength of the effect size. Also, cognitive processes tend to decline among samples of older adults (Urbina, 2004). Because research indicates a relative superiority on spatial navigation task of male participants (Parsons et al., 2004; Voyer, Voyer, & Bryden, 1995) we considered gender as a potential theoretically relevant moderator variable.

Type of clinical condition

No previous research has compared the performance of different clinical populations on the same cognitive process assessed by virtual reality-based measures. Thus, we aim to investigate whether the overall effect is more or less salient for particular types of clinical conditions.

Type of exploration within virtual reality environments

Two types of exploration can be identified within virtual scenarios. The first one is called ―active exploration‖. In this condition, the participants are immersed and navigate in the virtual environment. They are guided through the virtual world by a research assistant or navigate and move around by themselves with a joystick. They have a 360 ° ―first person‖ view of the

---

comparing two contrasted groups, usually coming from clinical and healthy populations on the performance obtained on specific cognitive functions (Lezak, 1995; Urbina, 2004; Wasserman & Bracken, 2003). After computing the magnitude of the difference between the performance of the two groups which is usually expressed in terms of effect size one can obtain estimates of test overlap and probability of superiority. It is considered that a diagnostic marker in neuropsychological assessment has an appropriate level of sensitivity if the test overlap is lower than 5% expressed in an effect size larger than 3.0 in magnitude according to Cohen's *d* metrics (Zakzanis, 2001).

environment. The second approach is ―the passive exploration‖. This time participants do not navigate in the virtual environment. They are immersed in the virtual world, but do not move around or explore it. They stay in a fixed location and are exposed to stimuli, but can look around and have a 360° ―first person‖ view of the environment. Consequently, we aim to investigate whether active or passive exploration can strengthen the overall effect.

Task performance indicator

We classified task performance indicator in three main clusters: (1) based on time, such as the reaction time, (2) based on errors, like correct or incorrect responses, and (3) the total amount of body movement recorded.

Presence of distractors

The presence of distractors can enhance the presence and the immersion into the virtual environment, as well as the ecological validity. Yet, some distractors may trigger more cognitive resources for the completion of a task, increasing task difficulty. As a consequence, we expected the presence or the absence of distractors to be a moderator variable.

Type of virtual reality platform

Two main types of virtual reality platforms are used in assessment or rehabilitation of cognitive processes. The first type of virtual reality platform is the Head Mounted Display (HMDs). This type of virtual reality platform provides a full 360° ―first person‖ view of the virtual environment while navigating. The projected video-capture virtual reality platform consists of a video camera that captures and converts the participant's movements in a 2D world on a large monitor. The participant sees himself in a mirror image. Previous research has indicated that these types of virtual reality platforms influence task performance (Rand et al., 2005).

Method

Literature search

We conducted a systematic literature search using ―virtual reality‖, ―cogn* assessment‖, ―memory‖, ―executive funct*‖, and ―attention‖ as search terms in Medline, PsycInfo and ScienceDirect databases, up to November 2014. Also, we screened reviews on the topic of virtual reality assessment and the list of references of empirical articles to detect studies that did not appear in the electronic search.

Studies selection

The following criteria were used for the inclusion of studies in the meta-analysis: (a) any experimental study with minimum two experimental groups: a healthy control group and a clinical group measured with the same virtual reality assessment tool; (b) there was sufficient data to compute effect sizes; (c) publications were in English.

The initial search procedure revealed 146 records plus 33 additional records identified through other sources (see Fig. 1). Sixteen duplicates were removed. A total of 163 potential abstracts were screened. Dissertations, publications in other languages than English, and studies that did not focus on virtual reality and cognitive assessment were not taken into account. A total of 115 potential articles were analyzed in detail based on their full text. Studies that used computer devices without immersion via HMDs or projected video-capture virtual reality platforms have been excluded. Eighteen studies met the inclusion criteria and were included in the meta-analysis.

Insert Figure 1 about here

Several studies were excluded because they used only pretest to posttest measures of a clinical group in the absence of a control group (Lee et al., 2003) or they used very small samples (only two participants) in the clinical group (Albani et al., 2002).

Data coding

The following variables were coded: study identification data, participants' mean age, percentage of male participants, number of participants per condition, type of clinical condition, type of exploration within virtual reality environments, task performance indicator, presence of distractors, type of immersion (HMD or projected video-capture platform), assessment duration within virtual reality environments, and type of cognitive process.

We classified the outcome measures into three categories based on the cognitive process assessed according to Lezak (1995), and subsequent cognitive assessment scales: executive functions[2], memory[3], and visuospatial analysis[4] measures, as they were the only available measures from the selected studies.

Effect size calculation and heterogeneity

We compared cognitive performance of the clinical group and control group measured with virtual reality-based assessment tools to address the sensitivity of virtual reality measures. In terms of Cohen's *d*,[5] superior performance of the healthy control group was considered as evidence for the sensitivity of virtual reality-based measures in detecting cognitive deficits.

---

[2]general measures of executive functioning, as well as impulsivity/inhibition and attention indexes/measures

[3]memory and learning processes (e.g., target recall, target recognition, total errors)

[4]spatial rotation and measures of visuospatial neglect

[5] accordingly to the specifications from the literature regarding the need to carefully interpret the magnitude of Cohen's *d* qualification of ―small‖, ―medium‖ and ―large‖ effects depending on the context (Zakzanis, 2001). Although, for instance, 1.0 is a large effect according to Cohen's metrics, this value reflects approximately only 45% percent overlap. This indicates that approximately 50% participants from the clinical group obtain scores different from those

For the first objective between-group effect sizes were calculated using Hedges's *g*[6]. In order to compute effect sizes, we used mean scores, standard deviations, and sample size. When there were studies that did not provide means and standard deviations we calculated *g* values from exact *t*, *F*, and *p* values applying conversion formulas when necessary. Thus, we obtained estimates of the effect and not the true effect as would be derived from means and standard deviations. Further, three studies had the mean and standard deviation equal to 0, meaning that there is no dispersion among results (Broeren et al., 2007; Kim et al., 2004; Parsons et al., 2007). As a consequence, we computed the results using one sample t test calculator (Soper, 2015) and obtained a *t* statistic value. We computed an average effect size for each study and used the study as the unit of analysis. For our comparison between clinical and control groups' cognitive performance on virtual-reality based measures, positive effect sizes were considered as in favor of healthy participants. The mean effect size was computed using random effects model which assumes two sources of variance: one is within study error, and second, variation in true effects across studies (Borenstein et al., 2009). To test for heterogeneity of the effect sizes, we considered two statistics: the homogeneity test Q and the I^2 index.

Then, we performed subgroup analysis for executive functions measures, memory measures, and visuospatial analysis measures. For executive functions, we used the random effect model because we had enough studies to include in the analysis. For memory and visuospatial analysis measures, we used the fixed effect model, given that there were few studies in each category (Borenstein et al., 2009). Although applying a random-effect meta-analysis is

---

obtained by healthy controls. Therefore, we will not only rely in our analyses on Cohen's *d* metrics, but will also provide estimates of test overlap to investigate the sensitivity of virtual reality measures and interpret the result through this frame

[6] a value of Hedge's *g* between 0.20 and 0.50 indicates a small effect, one between 0.50 and 0.80 indicates a medium effect, while a value larger than 0.80 indicates a large effect size (Cohen, 1988).

more realistic, it produces more generalizable results and is highly recommended since we expect between-studies variance due to a high heterogeneity across samples of populations; when dealing with a reduced number of studies the procedure is not recommended because the estimated between-studies variance is unreliable (Borenstein et al., 2009).

Publication bias

We used the Duval and Tweedie's trim-and-fill procedure to investigate publication bias (Duval & Tweedie, 2000). This procedure identifies studies with extreme effect sizes from one side of the funnel plot and re-computes the effect sizes taking into account hypothetical symmetrical counterparts of those extremes. The result is therefore an unbiased estimate of the effect size.

Software

The statistical analysis was conducted using Comprehensive Meta-Analysis software (version 2.2, Borenstein, Hedges, Higgins, & Rothstein, 2005).

Results

For the first objective, we computed average effect sizes from 18 studies comparing the performance of the clinical and control groups on virtual reality measures (N = 668). We found a large mean effect size in favor of the healthy control group (g = 0.95, 95% CI [0.67, 1.22], z = 6.75; $p < .001$). The percent overlap is 45% and there is a 76% chance, that a participant picked at random from the clinical group, to have a higher score than a participant picked random from the control group. Nevertheless, there was evidence of heterogeneity in the results ($Q_{(17)}$ = 46.50, $p < .001$; I^2 = 63.44%) which was addressed by performing the moderation analysis. Table 1 provides a synthetic view of the studies' characteristics. Figure 2 displays the forest plot and the effect size values with a 95% CI.

Insert Table 1 about here

Insert Figure 2 about here

The between-group analysis for executive functions measures computed on 13 studies (N = 474) revealed a medium to large effect size the control group showing higher performances than the clinical group (g = 0.77, 95% CI [0.55, 0.99], z = 6.11; p < .001) with a low level of heterogeneity ($Q_{(12)}$ = 6.86, p < .001; I^2 = 24.40%) and a percent overlap of 52% and a 71% probability of superiority. Also, between-group analysis for memory measures on three studies (N = 134) showed a mean overall significant effect size in favor of the control group (g = 0.96, 95% CI [0.59, 1.33], z = 5.09; p < .001), percent overlap of 45% and a 76% probability of superiority. Considering the increased heterogeneity ($Q_{(2)}$ = 22.37, p < .001; I^2 = 91.06%) and the fact that the analysis was conducted with only three studies, this result should be interpreted with caution. A third between-group analysis for visuospatial analysis measures was conducted on two studies (N= 60) and indicated significant differences between the control and clinical groups, with healthy participants from the control group having better results (g = 1.70, 95% CI [1.06, 2.34], z = 5.19; p < .001), ($Q_{(1)}$ = 0.00, p = .932; I^2 = .00%).

Moderation analysis

To investigate the second objective, we conducted moderation analysis.The results from the between-group analysis for performance on virtual reality cognitive measures revealed moderate variability in the effect sizes. As a consequence, a meta-regression analysis for numeric moderators and a subgroup analysis for categorical moderators were performed.

Results from the meta-regression show that age moderates the effect size (β = 0.005, 95% CI [-0.00, 0.00], z = 2.49; p < .05). The general tendency for the effect size is to increase together with the age of the participants. Nevertheless, due to the fact that the β coefficient has a value close to 0, the practical significance of the effect is null. Next, gender did not moderate the effect size of the performance on virtual reality-based measures.

The subgroup analysis revealed that the type of clinical condition moderated the mean weighted effect size with larger effects for brain injury (g = 0.90, 95% CI [0.78; 1.02], p < .001) and ADHD (g = 0.90, 95% CI [0.77; 1.02], p < .001), followed by schizophrenia (g = 0.78, 95% CI [0.54; 1.02], p < .001), and neurofibromatosis type 1 (g = 0.37, 95% CI [0.15; 0.59], p < .001). The type of exploration significantly moderated the effect size for the performance on virtual reality-based cognitive measure and the passive exploration (g = 0.89, 95% CI [0.79; 0.99], p < .05) outperformed the active exploration (g = 0.72, 95% CI [0.60; 0.84], p < .05). It seems that the difference between the clinical and non-clinical population is larger for the condition in which participants did not navigate throughout the virtual environment and were passively exposed to the stimuli compared that in which the participants explored and navigated through the environment. Another significant moderator is the presence of distractors with larger effect sizes for the no distractors condition (g = 0.94, 95% CI [0.82; 1.05], p < .01) compared to the virtual environment with distractors (g = 0.73, 95% CI [0.63; 0.83], p < .01). Moreover, the subgroup analysis revealed that the task performance indicator and the type of virtual reality platform do not moderate the effect size outcome, suggesting that they provided similar effect sizes and influenced equally the overall performance on virtual reality measures (see Table 2).

Insert Table 2 about here

Publication bias

For the comparison of clinical and healthy populations on virtual reality based measures, the trim-and-fill procedure identified no study with an effect higher than the mean which can modify the results. These analyses indicate that our results are robust and not affected by publication bias.

Discussion

The present meta-analysis investigated the sensitivity (as defined above) of virtual reality-based measures in detecting cognitive deficits by comparing performance of clinical population to healthy controls on several cognitive functions.

Overall, our findings provide support for the sensitivity of virtual reality-based assessment tools in detecting cognitive deficits. As expected, when we investigated the differences in performance on virtual reality-based measures between clinical and healthy populations on cognitive processes, we found that the healthy control group outperformed the clinical group (g = 0.95, 95% CI [0.67, 1.22]). These results are similar with those obtained in validation studies of different neuropsychological tests that aimed to discriminate between clinical and healthy population using the method of contrasted groups (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Frencham, Fox, & Maybery, 2005; Henry, Crawford, & Phillips, 2004). Because virtual reality-based measures showed significant differences between healthy individuals and patients with various conditions, we consider that virtual reality tests are sensitive in detecting cognitive impairment. However, although ~~according to Cohen's *d* benchmarks~~ our results point out to a large effect, if we analyze the sensitivity of the virtual reality-based measures by computing the percent overlap, its size is 45%, suggesting that almost half of the patients from the clinical group obtained different scores compared to non-clinical one. According to Zakzanis (2001), a diagnostic marker in neuropsychological testing should have an overlap lower than 5% and an effect size of at least 3.0. Based on this, we can say that the sensitivity of virtual reality measures in detecting cognitive deficits is moderate.

We also classified the outcome measures into the following categories based on the cognitive process assessed as recommended by Lezak (1995): executive functions measures, memory measures, and visuospatial analysis measures. We computed mean effect sizes and performed subgroup analysis for each of the three categories of cognitive processes to investigate any differences between the clinical and control group. For all categories, our data showed that the control group outperformed the clinical group, highlighting the sensitivity of virtual reality tests for different cognitive processes. Moreover, the magnitude of all effect sizes was medium to large. The largest mean effect size was in the case of visuospatial analysis measures (g = 1.70, 95% CI [1.06, 2.34], percent overlap 25%), followed by memory measures (g = 0.96, 95% CI [0.59, 1.33], percent overlap of 45%), and executive functions measures (g = 0.77, 95% CI [0.55, 0.99], percent overlap 53%). It is important to note that, for visuospatial analysis measures and memory measures only, two and respectively three studies were included in the analysis and inferences made from these results have limited reliability. Overall, these findings provide evidence that support a moderate level sensitivity of the virtual reality tests in detecting cognitive impairment.

We compared values of the effect size from the present study with those from meta-analyses that compared the performance of clinical versus healthy control groups on classical or computerized cognitive measures in order to discuss the sensitivity of virtual reality measures in comparison with traditional measures. The effect sizes of the performance of patients suffering from schizophrenia compared to healthy controls were medium to large, ranging from 0.85 to 1.21 for memory measures, 0.55 to 1.41 for executive functions, 0.71 to 0.96 for attention, and 0.91 for general cognitive ability (see Aleman, Hijman, de Haan, & Kahn, 1999; Bokat & Goldberg, 2003; Forbes, Carrick, McIntosh, & Lawrie, 2009; Henry & Crawford, 2004; Mesholam-Gately, Giuliano, Goff, Faraone, & Seidman, 2009). In the case of patients with brain injury, the comparisons revealed small to medium effect sizes. For example, for memory measures values ranged from 0.30 to 0.35, for executive functions from 0.15 to 0.48., for overall cognitive ability from 0.03 to 0.74, while for attention the effect size was 0.47 (see Belanger et al., 2005; Frencham et al., 2005; Henry et al., 2004; Rohling et al., 2011; Ruttan, Martin, Liu, Colella, & Green, 2008; Schretlen & Shapiro, 2003; Vanderploeg, Curtiss, & Belanger, 2005). In the case of patients with ADHD, the effect sizes had low to large values, depending on the type of cognitive process assessed. For memory measures, values ranged from 0.01 to 0.91, while for executive functions from 0.05 to 0.89, for attention, from 0.15 to 1.34, while for overall cognitive ability from 0.26 to 0.61 (see Boonstra, Oosterlaan, Sergeant, & Buitelaar, 2005; Bridgett & Walker, 2006; Frazier, Demaree, & Youngstrom, 2004; Hervey, Epstein, & Curry, 2004; Homack & Riccio, 2004; Huang-Pollock, Karalunas, Tam, & Moore, 2012; Lansbergen, Kenemans, & van Engeland, 2007; Losier, McGrath, & Klein, 1996). Overall, our results suggest a similar magnitude of the global effect size and sensitivity in detecting cognitive impairment for both virtual reality-based measures and classical paper-and-pencil and computerized psychometrics.

Moderator effects

Our second objective focused on moderator analysis because the main effect in the meta-analysis revealed heterogeneity.

We investigated age, gender, type of clinical condition, type of exploration, task performance indicator, the presence of distractors, and type of virtual reality platform as potential moderators of the differences in performance on virtual reality-based measures between clinical and healthy populations. Participants' age appeared to be a significant moderator. It seems that the more the age of the participants increases, the effect size increases. This could mean that in case of older participants virtual reality measures have an increased sensitivity compared to young participants. One could speculate that virtual reality-based measures target the cognitive decline associated with aging. The type of clinical condition is another significant moderator. Virtual reality-based measures have an increased sensitivity for brain injury or ADHD conditions followed by schizophrenia. For attention deficits associated with neurofibromatosis type 1, virtual reality measures seem to have limited sensitivity. However, for neurofibromatosis type 1, data was available from one study, so the reliability of the inferences made is limited. Another significant moderator that emerged was the type of exploration. Performance obtained in a passive exploration yielded a larger difference between clinical and healthy participants compared to active exploration. We speculate that active exploration might have an increased task difficulty and triggers more cognitive resources than passive exploration. Therefore, a task which implies an active exploration is difficult to both clinical and healthy participants, and the differences in performance between the two populations categories tend to reduce. This result is in line with the assumption regarding the increased level of task difficulty and complexity of assessment in virtual reality for clinical and healthy individuals (Armstrong et al., 2013; Elkind et al., 2001). Next, the presence or absence of distractors moderated the mean weighted effect size. The differences in performance between the clinical and control group are larger for the virtual reality measures that did not include distractors in the scenario. It seems that the virtual reality measures discriminate better between controls and clinical populations if the environments do not include distractors. The other potential moderators did not alter the overall effect size. Because the task performance indicator did not moderate the overall effect size, we consider these results as evidence for the effectiveness of virtual reality-based measures for time based, error based and body movement measures. We expected the type of virtual reality platform to be a significant moderator because previous research has identified differences in performance between the HMDs or gesture-based video-capture systems (Rand et al., 2009). Such a result points out that irrespective of platform type, virtual reality-based measure discriminate between healthy or controls.

Limitations and conclusions

One shortcoming of our research effort is the small number of studies that were included in the main analysis, as well as in the subgroup analysis, as it may weaken the statistical power. In some cases, the subgroup analysis was performed with a small number of studies for each category, restricting the robustness and reliability of the analysis. Also, while the spirit of a meta-analysis is to include all possible studies is an aspirational one, it is quite probable that not all studies were included that would have met inclusion criteria despite our increased effort to include all eligible studies.

Although the current meta-analysis brings evidence for the sensitivity of virtual reality measures by providing mean weighted effect size, future research should focus on more reliable indexes of diagnostic validity, such as sensitivity, specificity, positive predictive power, and negative predictive power. Few studies have investigated the predictive validity of virtual reality-based measures in relationship to real-life performance or other objective criteria. There is a need for studies to verify the predictive validity of virtual reality-based measures. Other studies might seek to set norms and to perform reliability analysis for virtual reality-based measures. Although providing normative data and performing reliability analysis is central to psychological testing (Urbina, 2004), to our knowledge, only one virtual reality-based measure designed to measure attention impairments in children with ADHD, AULA virtual reality test (Díaz-Orueta et al., 2014; Iriarte et al., 2012) is standardized.

In conclusion, our analysis supports the use of virtual reality-based measures in the neuropsychological assessment, because they are sensitive in detecting abnormal cognitive functioning. Having medium to large effects for each cognitive process, researchers and clinicians might use virtual reality measures to target cognitive deficits. However, it is very important to notice that when looking at the level of sensitivity, the virtual reality measures show a moderate level of sensitivity in correctly detecting cognitive deficits in patients.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. A. (2009). Distractibility in Attention/Deficit/ Hyperactivity Disorder (ADHD): the Virtual reality classroom. *Child Neuropsychology, 15*(2), 120-135. doi: 10.1080/09297040802169077

Albani, G., Pignatti, R., Bertella, L., Priano, L., Semenza, C., Molinari, E., & Mauro, A. (2002). Common daily activities in the virtual environment: a preliminary study in parkinsonian patients. *Neurological Sciences, 23*(2), s49-s50. doi: 10.1007/s100720200064

Aleman, A., Hijman, R., de Haan, E. H., & Kahn, R. S. (1999). Memory impairment in schizophrenia: a meta-analysis. *American Journal of Psychiatry, 156*(9), 1358-1366. doi: psychiatryonline.org/doi/pdf/10.1176/ajp.156.9.1358

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A Meta-analytic review. *Neuropsychology Review, 16*(1), 17-42. doi: 10.1007/s11065-006-9002-x

Armstrong, C. M., Reger, G. M., Edwards, J., Rizzo, A. A., Courtney, C. G., & Parsons, T. D. (2013). Validity of the Virtual Reality Stroop Task (VRST) in active duty military. *Journal of Clinical and Experimental Neuropsychology, 35*(2), 113-123. doi:10.1080/13803395.2012.740002

*Banville, F., Nolin, P., Lalonde, S., Henry, M., Dery, M. P., & Villemure, R. (2010). Multitasking and prospective memory: Can virtual reality be useful for diagnosis? *Behavioural Neurology, 23*(4), 209–211. doi: 10.3233/BEN20100297

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215-227. doi: 10.1017/S1355617705050277

* Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., & Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *European Journal of Paediatric Neurology, 16*, 514-521. doi: 10.1016/j.ejpn.2012.01.006

Bokat, C. E., & Goldberg, T. E. (2003). Letter and category fluency in schizophrenic patients: a meta-analysis. *Schizophrenia \research, 64*(1), 73-78. doi: 10.1016/S0920-9964(02)00282-7

Boonstra, A. M., Oosterlaan, J., Sergeant, J. A., & Buitelaar, J. K. (2005). Executive functioning in adult ADHD: a meta-analytic review. *Psychological Medicine, 35*(8), 1097-1108. doi: 10.1017/S003329170500499X

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2005). Comprehensive meta-analysis version 2. Englewood, NJ: Biostat

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: John Wiley & Sons.

Bridgett, D. J., & Walker, M. E. (2006). Intellectual functioning in adults with ADHD: a meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological Assessment, 18*(1), 1-14. doi: 10.1037/1040-3590.18.1.1

*Broeren, J., Samuelsson, H., Stibrant- Sunnerhagen, K., Blomstrand, C., & Rydmark, M. (2007). Neglect assessment as an application of virtual reality. *Acta Neurologica Scandinavica, 116*(3), 157-163. doi: 10.1111/j.1600-0404.2007.00821.x

Chan, C. L. F., Ngai, E. K. Y., Leung, P. C. H., & Wong, S. (2010). Effect of the adapted Virtual reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. *International Journal of Geriatric Psychiatry, 25*(6), 643-649. doi: 10.1002/gps.2403

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A Review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197. doi: 10.1023/B:NERV.0000009483.91468.fb

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, New Jersey: La Wrence Erlbaum Associates, Publishers.

Díaz-Orueta, U., Alvarado, S., Gutiérrez, D., Climent, G., & Banterla, F. (2012). ―Isla Calma‖, a novel virtual reality environment for pain and anxiety distraction: Report on usability, acceptability, and subjective experience. *GAMES FOR HEALTH JOURNAL: Research, Development, and Clinical Applications, 1*(5), 353-361. doi: 10.1089/g4h.2012.0037

Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G., & Narbona, J. (2014). AULA virtual reality test as an attention measure: Convergent validity with Conners' Continuous Performance Test. *Child Neuropsychology, 20*(3), 328-342. doi: 10.1080/09297049.2013.792332

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455– 463. doi: 10.1111/j.0006-341X.2000.00455.x

Elkind, J. S. (1998). Use of virtual reality to diagnose and habilitate people with neurological dysfunctions. *CyberPsychology & Behavior, 1*(3), 263-273. doi: 10.1089/cpb.1998.1.263

Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of preliminary results. *CyberPsychology & Behavior, 4*(4), 489-496. doi: 10.1089/109493101750527042

*Erez, N., Weiss, P. L., Kizony, R., & Rand, D. (2013). Comparing performance within a virtual supermarket of children with traumatic brain injury to typically developing children: A pilot study. *OTJR: Occupation, Participation & Health, 33*(44), 218-227. doi: 10.3928/15394492-20130912-04

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychological Medicine, 39*(6), 889-905. doi: 10.1017/S0033291708004558

Foreman, N., & Stirk, J. (2005). Commentary on Rose, F.D., Brooks, B.M., & Rizzo, A.A., virtual reality in brain damage rehabilitation: Review. *CyberPsychology & Behavior, 8*(3), 263-264. doi: 10.1089/cpb.2005.8.263

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology, 18*(3), 543-555. doi: 10.1037/0894-4105.18.3.543

Frencham, K. A., Fox, A. M., & Maybery, M. T. (2005). Neuropsychological studies of mild traumatic brain injury: A meta-analytic review of research since 1995. *Journal of Clinical and Experimental Neuropsychology*, *27*(3), 334-351. doi: 10.1080/13803390490520328

Gamberini, L. (2000). Virtual reality as a new research tool for the study of human memory. *CyberPsychology & Behavior, 3*(3), 337-342. doi: 10.1089/10949310050078779

*Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadef, H., Rizzo, A. S., & Josman, N. (2011). Using a Virtual Classroom environment to describe the attention deficits profile of children with Neurofibromatosis type 1. *Research in Developmental Disabilities, 32*(6), 2608–2613. doi: 10.1016/j.ridd.2011.06.014

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology Review, 18*(4), 621-628. doi: 10.1037/0894-4105.18.4.621

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia, 42*(9), 1212-1222. doi: 10.1016/j.neuropsychologia.2004.02.001

Henry, M., Joyal, C. C., & Nolin, P. (2012). Development and initial assessment of a new paradigm for assessing cognitive and motor inhibition: The bimodal virtual-reality Stroop. *Journal of Neuroscience Methods, 210*(2), 125-131. doi: 10.1016/j.jneumeth.2012.07.025.

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: a meta-analytic review. *Neuropsychology, 18*(3), 485-503. doi: 10.1037/0894-4105.18.3.485

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology, 19*(6), 725-743. doi: 10.1016/j.acn.2003.09.003

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: a meta-analysis of CPT performance. *Journal of Abnormal Psychology, 121*(2), 360-371. doi: 10.1037/a0027205

Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., & Climent, G. (2012). AULA—Advanced virtual reality tool for the assessment of attention: normative study in Spain. *Journal of Attention Disorders*, 1-27. doi: 10.1177/1087054712465335

*Kang, Y. J., Ku, J., Han, K., Kim, S. I., Yu, T. W., Lee, J. H., & Park, C. I. (2008). Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. *CyberPsychology & Behavior, 11*(3), 329-339. doi: 10.1089/cpb.2007.0116

* Kim, K., Kim, J., Ku, J., Kim, D. Y., Chang, W. H., Shin, D. I., ... & Kim, S. I. (2004). A virtual reality assessment and training system for unilateral neglect. *CyberPsychology & Behavior, 7*(6), 742-749. doi: 10.1089/cpb.2004.7.742

*Ku, J., Cho, W. G., Kim, J. H., Kim, K. U., Kim, B. N., Hahn, W. Y., ... Kim, S. I. (2004). The development of a VR system for the cognitive & behavioral assessment of schizophrenia. In J. D. Westwood et al. (Eds.), *Studies in Health Technology and Informatics* (pp. 180-182). Amsterdam: IOS Press. doi: 10.3233/978-1-60750-942-4-180

*Ku, J., Cho, W., Kim, J. J., Peled, A., Wiederhold, B. K., Wiederhold, M. D., … & Kim, S. I. (2003). A virtual environment for investigating schizophrenic patients' characteristics: assessment of cognitive and navigation ability. *CyberPsychology & Behavior, 6*(4), 397-404. doi: 10.1089/109493103322278781

Lalonde, G., Henry, M., Drouin-Germain, A., Nolin, P., & Beauchamp, M. H. (2013). Assessment of executive function in adolescence: a comparison of traditional and virtual reality tools. *Journal of Neuroscience Methods, 219*(1), 76-82. doi: 10.1016/j.jneumeth.2013.07.00

Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: a review and meta-analysis. *Neuropsychology, 21*(2), 251-262. doi: 10.1037/0894-4105.21.2.251

Lee, J. H., Ku, J., Cho, W., Hahn, W. Y., Kim, I. Y., Lee, S. M., & Kim, S. I. (2003). A virtual reality system for the assessment and rehabilitation of the activities of daily living. *CyberPsychology & Behavior, 6*(4), 383-388. doi: 10.1089/109493103322278763

Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). New York: Oxford University Press.

Lloyd, J., Riley, G. A., & Powell, T. E. (2009). Errorless learning of novel routes through a virtual town in people with acquired brain injury. *Neuropsychological Rehabilitation, 19*(1), 98-109. doi: 10.1080/09602010802117392

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the continuous performance test in non- medicated and medicated samples of children with and without adhd: A meta- analytic review. *Journal of Child Psychology and Psychiatry, 37*(8), 971-987. doi: 10.1111/j.1469-7610.1996.tb01494.x

*Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, S. R., Millis, S. R., & S., R. A. (2007). Is learning and memory different in a virtual environment? *The Clinical Neuropsychologist, 21*(1), 146-161. doi: 10.1080/13854040601100668

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. . *Neuropsychology, 23*(3), 315-336. doi: 10.1037/a0014708

*Moreau, G., Guay, M.-C., Achim, A., Rizzo, A., & Lageix, P. (2006). The virtual classroom: An ecological version of the continuous performance test - A pilot study. In

B. C. Wiederhold , S. Bouchard & G. Riva (Eds.), *Annual Review of Cybertherapy and Medicine* (Vol. 4, pp. 59-65). San Diego, CA: Interactive Media Institute.

Morganti, F. (2004). Virtual interaction in cognitive neuropsychology. *Studies in Health Technology and Informatics, 99*, 55-70. doi: 10.3233/978-1-60750-943-1-55

Myers, R. L., & Bierig, T. A. (2000). Virtual reality and left hemineglect: a technology for assessment and therapy. *CyberPsychology & Behavior, 3*(3), 465-468. doi: 10.1089/10949310050078922

Opriş, D., Pintea, S., Garcia-Palacios, A., Botella, C., Szamoskozi, S., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depress Anxiety, 29*(2), 85-93. doi: 10.1002/da.20910

Parsons, T. D., Courtney, C. G., & Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *Journal of Clinical and Experimental Neuropsychology, 35*(8), 812-826. doi: 10.1080/13803395.2013.824556

Parsons, T. D., Courtney, C. G., Arizmendi, B. J., & Dawson, M. E. . (2011). Virtual reality stroop task for neurocognitive assessment. In J. D. Westwood (Ed.), *Medicine Meets Virtual Reality* (pp. 433-439). Amsterdam: IOS Press.

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42*(4), 555-562. doi: 10.1016/j.neuropsychologia.2003.08.014

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy And Experimental Psychiatry, 39*(3), 250-261. doi: 10.1016/j.jbtep.2007.07.007

Parsons, T. D., & Rizzo, A. A. (2008). Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. *CyberPsychology & Behavior, 11*(1), 17-25. doi: 10.1089/cpb.2007.9934

*Parsons, T. D., Rizzo, A. A., Bamattre, J., & Brennan, J. (2007). Virtual reality cognitive performance assessment test. *Annual Review of CyberTherapy and Telemedicine*, 143-149.

*Pollak, Y., Weiss, P. L., Rizzo, A. A., Weizer, M., Shriki, L., Shalev, R. S., & Gross-Tsur, V. (2009). The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *Journal of Developmental & Behavioral Pediatrics, 30*(1), 2-6. doi: 10.1097/DBP.0b013e3181969b22

Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, *22*(3), 561-569. doi: 10.1016/j.janxdis.2007.04.006

Pugnetti, L., Attree, E. A., Mendozzi, L., Barbieri, E., Motta, A., Alpini, D., & Rose, F. D. (1999). Developments of collaborative research on VR applications in mental health. Focus on cybersickness and memory testing. *The International Journal of Virtual Reality, 4*(1), 41-47.

*Pugnetti, L., Mendozzi, L., Attree, E. A., Barbieri, E., Brooks, B. M., & Cazzullo, C. L. (1998a). Probing memory and executive functions with virtual reality: Past and present studies. *CyberPsychology & Behavior, 1*(2), 151–162. doi: 10.1089/cpb.1998.1.151

Pugnetti, L., Mendozzi, L., Barbieri, E., & Motta, A. (1998b). VR Experience with Neurological Patients: Basic Cost/Benefit Issues. In G. Riva, B. C. Wiederhold & E.

Molinari (Eds.), *Virtual Environments in Clinical Psychology and Neuroscience* (pp. 243-248). Amsterdam: IOS Press.

*Rand, D., Basha-Abu Rukan , S., Weiss, P. L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation: An International Journal of Head Trauma Rehabilitation, 19*(4), 583-602. doi: 10.1080/09602010802469074

*Rand, D., Katz, N., & Weiss, P. L. (2007). Evaluation of virtual shopping in the VMall: comparison of post-stroke participants to healthy control groups. *Disability and Rehabilitation*, *29*(22), 1710-1719. doi: 10.1080/09638280601107450

Rand, D., Kizony, R., Feintuch, U., Katz, N., Josman, N., Rizzo, A., & Weiss, P. L. (2005). Comparison of two VR platforms for rehabilitation: Video capture versus HMD. *Presence*, *14*(2), 147–160. doi: 10.1162/1054746053967012

Rheingold, H. (1991). *Virtual Reality*. New York: Summit.

Riva, G. (1998). Virtual Reality as Assessment Tool in Psychology. In G. Riva (Ed.), *Virtual Reality in Neuro-Psycho-Physiology* (pp. 71-79). Amsterdam: IOS Press.

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: the Virtual Classroom. *CNS Spectrums, 11*(1), 35-44. doi: 10.1017/S1092852900024196

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., & Sisemore, D. (2000). The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychology & Behavior, 3*(3), 483-499. doi: 10.1089/10949310050078940

Rizzo, A. A., Buckwalter, J. G., Neumann, U., Chua, C., Van Rooyen, A., Larson, P., & Humphrey, L. (1999). Virtual environments for targeting cognitive processes: an overview of projects at the University of Southern California. *CyberPsychology & Behavior, 2*(2), 89-100. doi: 10.1089/cpb.1999.2.89

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation, 14*(1-2), 207-239. doi: 10.1080/09602010343000183

Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009). *The Clinical Neuropsychologist, 25*(4), 608-623. doi: 10.1080/13854046.2011.565076

Rose, F. D., Brooks, B. M., & Rizzo, A. A. (2005). Virtual reality in brain damage rehabilitation: Review. *CyberPsychology & Behavior, 8*(3), 241-262. doi: 10.1089/cpb.2005.8.241

Ruttan, L., Martin, K., Liu, A., Colella, B., & Green, R. E. (2008). Long-term cognitive outcome in moderate to severe traumatic brain injury: a meta-analysis examining timed and untimed tests at 1 and 4.5 or more years after injury. *Archives of Physical Medicine and Rehabilitation, 89*(12), S69-S76. doi: 10.1016/j.apmr.2008.07.007

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15*(4), 341-349. doi: 10.1080/09540260310001606728

Schultheis, M. T., Himelstein, J., & Rizzo, A. A. (2002). Virtual reality and neuropsychology: upgrading the current tools. *Journal of Head Trauma Rehabilitation, 17*(5), 378-394.

*Siemerkus, J., Irle, E., Schmidt-Samoa, C., Dechent, P., & Weniger, G. (2012). Egocentric spatial learning in schizophrenia investigated with functional magnetic resonance imaging. *NeuroImage: Clinical, 1*(1), 153-163. doi: 10.1016/j.nicl.2012.10.004

Soper, D. (2015). Statistics Calculators. Retrieved from http://www.danielsoper.com/statcalc3/calc.aspx?id=98

Sorita, E., N'Kaoua, B., Larrue, F., Criquillon, J., Simion, A., Sauzeon, H., . . . Mazaux, J. M. (2013). Do patients with traumatic brain injury learn a route in the same way in real and virtual environments? *Disability and Rehabilitation, 35*(16), 1371-1379. doi: 10.3109/09638288.2012.738761

Urbina, S. (2004). *Essentials of Psychological Testing*. New Jersey: John Wiley & Sons.

Vanderploeg, R. D., Curtiss, G., & Belanger, H. G. (2005). Long-term neuropsychological outcomes following mild traumatic brain injury. *Journal of the International Neuropsychological Society, 11*(2), 228-236. doi: 10.10170S1355617705050289

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270. doi: 10.1037/0033-2909.117.2.250

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of Psychology* (pp. 43-66). New Jersey: John Wiley & Sons.

Zakzanis, K. K. (2001). Statistics to tell the truth, the whole truth, and nothing but the truth: formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology*, *16*(7), 653-667. doi: 10.1016/S0887-6177(00)00076-7

Table 1

Characteristics of the studies included in the meta-analysis

Author(s)	Mean age (years)	% of male participants	*N*	Type of exploration	Task performance indicator	Distractors	Type of clinical condition	Type of cognitive process assessed	Outcome measure	Type of VR platform	Effect size (Hedges’s *g*)
Adams, Finn, Moes, Flannery, and Rizzo (2009)	10.30	100	34	Passive	Error-based measures	Yes	ADHD	Executive functions	VR Classroom Commissions, correct percent, cued recall, omissions	HMD	0.59
Banville, Nolin, Lalonde, Henry, Dery, and Villemure, 2010	27	74.19	62	Active	Error-based measures, Time-based measures	No	Brain injury	Memory	VR memory task precision, time to complete, succes in task	HMD	0.52
Bioulac, Lallemand, Rizzo, Philip, Fabrigoule, and Bouvard, 2012	8.28	100	36	Passive	Error-based measures, Time-based measures	Yes	ADHD	Executive functions	VR Classroom Total correct hits, commissions, correct hits reaction time, reaction time variability, commissions reaction time	HMD	0.26
Broeren, Samuelsson, Stibrant-Sunnerhagen,	54.37	50	8	Active	Error-based measures	No	Brain injury	Visuospatial analysis measures.	VR task Visuospatial neglect	HMD	1.65

Blomstrand, and Rydmark (2007)											
Erez, Weiss, Kizony, and Rand, 2013	12.40	50	40	Active	Error-based measures, Time-based measures	Yes	Brain injury	Executive functions	VMall time, number of mistakes	Gesture-based video-capture system	0.81
Gilboa, Rosenblum, Fattal-Valevski, Toledano-Alhadef, Rizzo, and Josman, 2011	12.20	29.62	54	Passive	Error-based measures, Time-based measures, Body movement	Yes	Neurofibromatosis type 1	Executive functions	VR Classroom Total correct hits, commissions, hits reaction type, head movement	HMD	0.38
Kang, Jeonghun, Han, Kim, Yu, Lee, and Park (2008)	52.95	62.50	40	Active	Error-based measures	No	Brain injury	Executive functions	VR SS Attention index, executive index, performance index	HMD	1.48
Kim, Kim, Ku, Kim, Chang, Shin, Lee, Kim, and Kim (2004)	48.10	53.84	52	Passive	Error-based measures, Time-based measures	No	Brain injury	Visuospatial analysis measures.	VR task Deviation angle, failure rate, no-attention time, number of cues, ratio of scan, scanning time	HMD	1.71
Ku, Cho, Kim,	25.46*	79.80*	33	Active	Error-based	Yes	Schizophre	Executive	VR Environment	HMD	1.00

Kim, Kim, Hahn, Kim, Lee, and Kim (2004)					measures		nia	functions	Perseverance index, rule finding index		
Ku, Cho, Kim,	28.95	53.84	26	Active	Error-based	Yes	Schizophre	Executive	VR Environment	HMD	1.00
Peled, Wiederhold, Wiederhold, Kim, Lee, and Kim (2003)					measures		nia	functions	Perseverance index, rule finding index		
Matheis, Schultheis, Tiersky, DeLuca, Millis, and Rizzo (2007)	36.23	50	40	Passive	Error-based measures	No	Brain injury	Memory	VR Office Memory recall, memory recognition	HMD	2.95
Moreau, Guay, Achim, Rizzo, and Lageix, 2006	25.46*	100	22	Passive	Error-based measures, Time-based measures	Yes	ADHD	Executive functions	VR Classroom Omissions, reaction time variability,	HMD	1.09
Parsons, Bowerly, Buckwalter, and Rizzo (2007)	10.40	100	20	Passive	Error-based measures, Time-based measures, Body movement	Yes, No	ADHD	Executive functions	VR Classroom Body movement, commissions, hit reaction time, omissions	HMD	1.20
Pollak, Weiss, Rizzo,	12.60	100	37	Passive	Error-based measures,	Yes	ADHD	Executive functions	VR Classroom Reaction time,	HMD	0.98

Weizer, Shriki, Shalev, and Gross-Tsur, 2009					Time-based measures				variability of reaction time, omissions, commissions		
Pugnetti, Mendozzi, Attree, Barbieri, Brooks, Cazzullo, Motta, and Rose (1998)	36.50	79.80*	63	Active	Error-based measures	No	Brain injury	Executive functions	VR WCST Categories achieved, executive function, total correct, total errors	HMD	0.46
Rand, Katz, and Weiss, 2007	60.50	62.50	40	Active	Error-based measures, Time-based measures	Yes	Brain injury	Executive functions	VMall time to complete the task, number of products bought by mistake	Gesture-based video-capture system	0.30
Rand, Rukan, Weiss, and Katz, 2009	64.10	58.62	29	Active	Error-based measures	Yes	Brain injury	Executive functions	VMall total mistakes, mistakes in completing tasks, partial mistakes, non efficiency mistakes, rule break mistakes, use of strategies mistakes	Gesture-based video-capture system	1.33
Siemerkus, Irle, Schmidt-Samoa, Dechent, and Weniger (2012)	27.90	65.62	32	Active	Error-based measures, Time-based measures	No	Schizophrenia	Memory	VR maze Repetitive errors, successful trials, total errors, total time	HMD	0.64

Note. Total *N* = 668; VMall = Virtual Mall (Rand, Katz, Shahar, Kizony, & Weiss, 2005); VR Classroom = Virtual Classroom (A. A. Rizzo et al., 2000); VR Environment = Virtual Reality Environment (Ku et al., 2003); VR maze = Virtual Reality Environment replicating a maze (Siemerkus, Irle, Schmidt-Samoa, Dechent, & Weniger, 2012); VR memory = Virtual Reality prospective memory (Banville et al., 2010); VR Office = Virtual Reality Office (Matheis et al., 2007); VR SS = Virtual Reality Shopping Simulation (Kang et al., 2008); VR task = Virtual environment to assess unilateral neglect (Kim et al., 2004); VR task visuospatial neglect = Cancellation test developed in the Virtual reality environment (Broeren et al., 2007); VR WCST = Virtual Reality analog of Wisconsin Card Sorting Test (Pugnetti, Mendozzi, Attree, et al., 1998) * = Mean age and mean of % of male participants were not provided in the studies and were substitute with the non-missing mean age and mean of % percentage of male participants of the studies included in the meta-analysis

Table 2

Moderation analysis with categorical variables for performance on virtual reality cognitive measures

Outcome	Moderator	*K*	*g*	*p*	Q w	*p*	*95%* CI	Q b	*p*
Performance on cognitive measures	Schizophrenia/	3	0.78	.000	3.41	.906	[0.54; 1.02]	19.37	.000
	brain injury/	9	0.90	.000	200.19	.000	[0.78; 1.02]		
	ADHD/	5	0.90	.000	78.15	.000	[0.77; 1.02]		
	Neurofibromatosis type 1	1	0.37	.001	6.25	.282	[0.15; 0.59]		
	Active exploration/	10	0.72	.000	99.53	.000	[0.60; 0.84]	4.67	.031
	Passive exploration	8	0.89	.000	203.20	.000	[0.79; 0.99]		
	Time-based measures/	10	0.73	.000	66.55	.000	[0.57; 0.89]	2.78	.248
	Error-based measures/	18	0.82	.000	196.22	.0.00	[0.72; 0.92]		
	Body movement	2	0.93	.000	41.84	.000	[0.75; 1.11]		
	Distractors/	11	0.73	.000	149.06	.000	[0.63; 0.83]	6.99	.008
	No distractors	8	0.94	.000	151.35	.000	[0.82; 1.05]		
	HMD/	15	0.83	.000	247.03	.000	[0.75; 0.91]	0.43	.509
	Gesture-based video capture system	3	0.75	.000	59.93	.000	[0.53; 0.97]		

Note. *K* = number of studies included in the analysis; *g* = Hedge's *g*; 95% CI = 95% confidence interval around the weighted mean effect size.

Virtual reality measures in neuropsychological assessment: a meta-analytic review

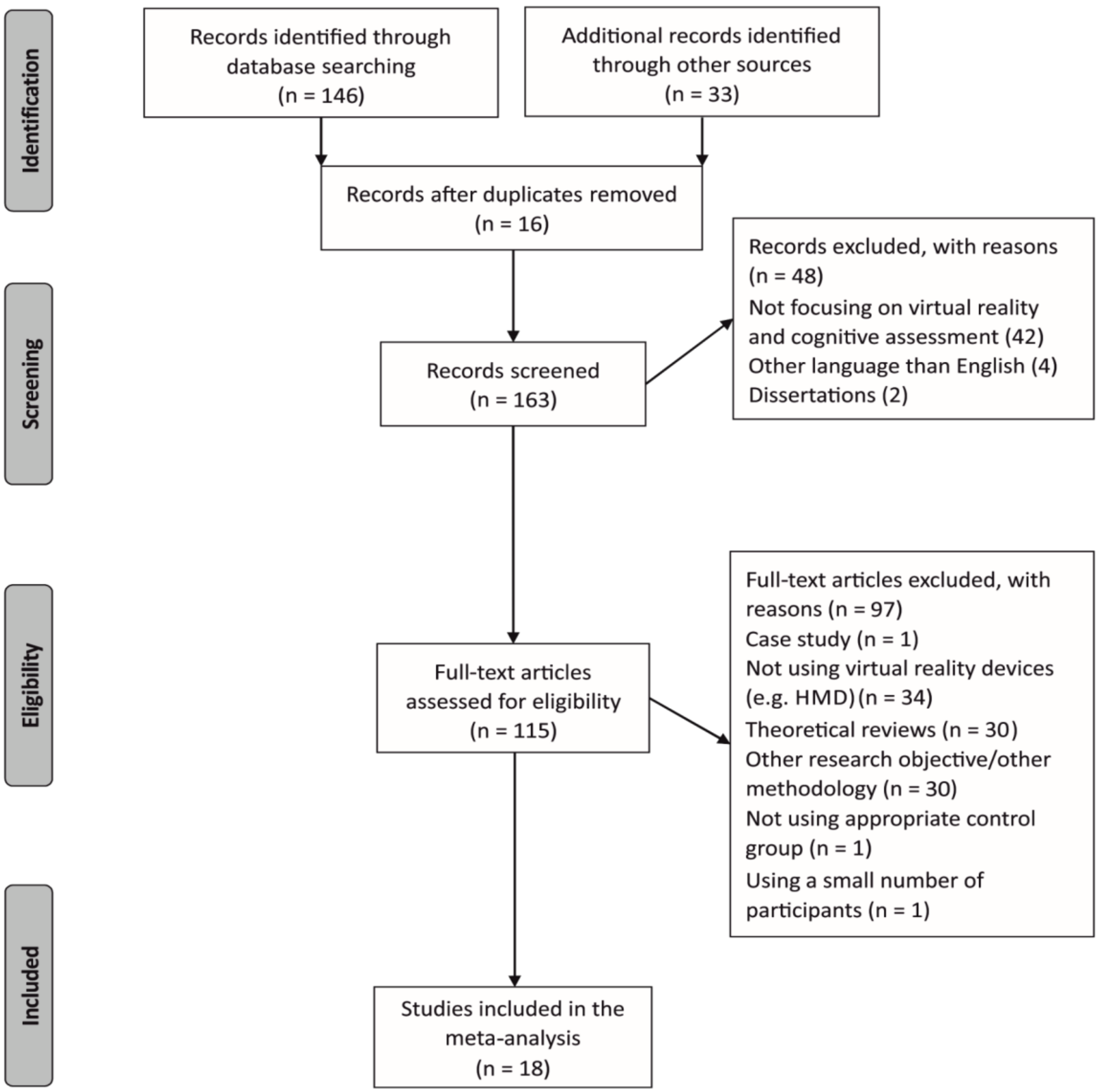

Figure 1. PRISMA flow diagram

Comparison of the performance between clinical and control group on virtual reality-based measures

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Adams et al., 2009	0.592	0.346	0.120	-0.087	1.270	1.709	0.088
Banville et al., 2010	0.521	0.256	0.065	0.020	1.022	2.037	0.042
Bioulac et al., 2012	0.268	0.330	0.109	-0.380	0.915	0.810	0.418
Broeren et al., 2007	1.651	0.688	0.473	0.303	2.999	2.401	0.016
Erez et al., 2013	0.812	0.323	0.105	0.178	1.446	2.510	0.012
Gilboa et al., 2011	0.385	0.273	0.074	-0.150	0.920	1.412	0.158
Kang et al., 2008	1.485	0.352	0.124	0.795	2.174	4.221	0.000
Kim et al., 2004	1.718	0.373	0.139	0.987	2.449	4.605	0.000
Ku et al., 2003	1.000	0.404	0.164	0.207	1.792	2.472	0.013
Ku et al., 2004	1.003	0.363	0.132	0.292	1.714	2.765	0.006
Matheis et al., 2007	2.954	0.462	0.214	2.049	3.860	6.393	0.000
Moreau et al., 2006	1.091	0.471	0.222	0.167	2.014	2.315	0.021
Parsons et al., 2007a	1.200	0.472	0.223	0.275	2.125	2.541	0.011
Pollak et al., 2009	0.982	0.345	0.119	0.305	1.658	2.845	0.004
Pugnetti et al., 1998	0.467	0.256	0.066	-0.035	0.969	1.823	0.068
Rand et al., 2009	1.330	0.441	0.194	0.466	2.194	3.017	0.003
Rand et el., 2007	0.305	0.347	0.120	-0.375	0.985	0.879	0.379
Siemerkus et al., 2012	0.644	0.355	0.126	-0.051	1.339	1.815	0.069
	0.951	0.141	0.020	0.675	1.226	6.756	0.000

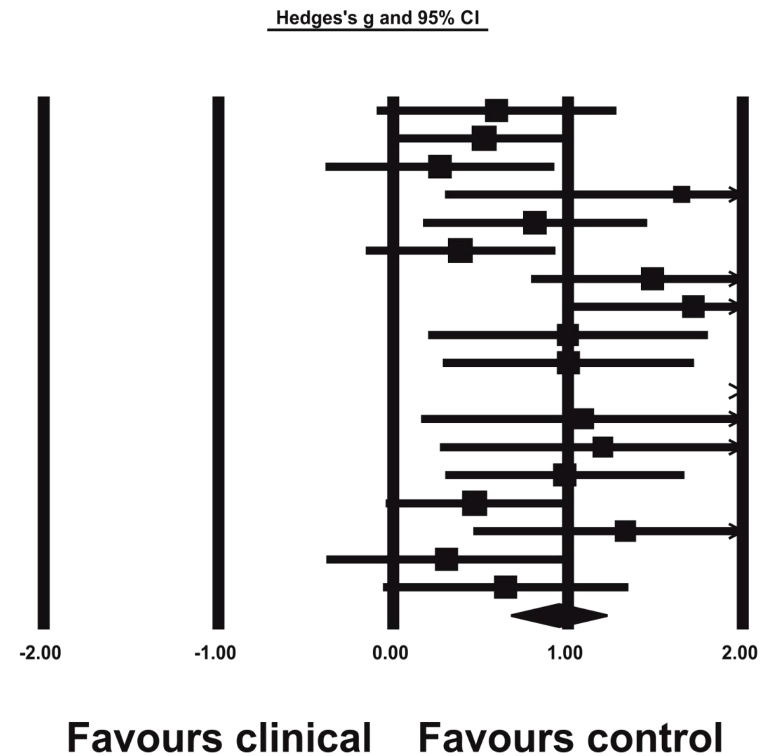

Figure 2. Meta-analysis for comparison of the performance between clinical and control group on virtual-reality-based measures